

THE DIFFERENTIAL OF THE EXPONENTIAL MAP, JACOBI FIELDS AND EXACT PRINCIPAL GEODESIC ANALYSIS *

S. SOMMER^{*†}, F. LAUZE[†], AND M. NIELSEN^{†‡}

Abstract. The importance of manifolds and Riemannian geometry is spreading to applied fields in which the need to model non-linear structure has spurred wide-spread interest in geometry. The transfer of interest has created demand for methods for computing classical constructs of geometry on manifolds occurring in practical applications. This paper develops initial value problems for the computation of the differential of the exponential map and Jacobi fields on parametrically and implicitly represented manifolds. It is shown how the solution to these problems allow for determining sectional curvatures and provides upper bounds for injectivity radii. In addition, when combined with the second derivative of the exponential map, the initial value problems allow for numerical computation of Principal Geodesic Analysis, a non-linear version of the Principal Component Analysis procedure for estimating variability in datasets. The paper develops algorithms for computing Principal Geodesic Analysis without the tangent space approximation previously used and, thereby, provides an example of how the constructs of theoretical geometry apply to solving problems in statistics. By testing the algorithms on synthetic datasets, we show how curvature affects the result of PGA.

Key words. manifolds, Riemannian metrics, computations on manifolds, manifold valued statistics, principal component analysis (PCA), principal geodesic analysis (PGA), geodesic principal component analysis (GPCA), tangent space linearization

1. Introduction. Manifolds, sets locally modeled by Euclidean spaces, have a long and intriguing history in mathematics, and topological, differential geometric, and Riemannian geometric properties of manifolds have been studied extensively with results extending far beyond the fields of manifolds themselves. The introduction of high-performance computing in applied fields have widened the use of manifolds, and Riemannian manifolds, in particular, are now used for modeling a range of problems possessing non-linear structure. Applications include shape modeling (complex projective shape spaces [23] and medial representations of surfaces [1, 20]), imaging (tensor manifolds in diffusion tensor imaging [11, 12, 31] and image segmentation and registration [4, 32]), and several other fields (forestry [18], human motion modeling [37, 27, 40]).

To fully utilize the power of manifolds in modeling, it is essential to develop fast and robust algorithms for computing various manifold constructions. Computing intrinsic distances, Jacobi fields, curvatures, and injectivity radii pose important problems [18] as well as solving optimization problems posed on manifolds or in manifold tangent spaces and defining and computing manifold generalizations of common Euclidean space statistics. The papers [6, 22, 29, 24, 36, 39] address first-order manifold problems, and certain second-order problems have been considered but mainly on limited classes of manifolds [8]. Generalizing linear statistics has been the focus of the papers [21, 30, 13, 10, 18].

In this article, we study the second-order problems arising from variations of the initial velocity of geodesics. This will allow us to compute structures fundamental to geometry and to numerically solve certain optimization problems posed in tangent

* The work is part of the LIMB project (Learning Imaging Biomarkers), a joint initiative of the Image Group, Department of Computer Science, University of Copenhagen and Nordic Bioscience Imaging A/S, Herlev, Denmark.

* (✉) sommer@diku.dk

†Department of Computer Science, University of Copenhagen, Denmark

‡Nordic Bioscience Imaging, Herlev, Denmark

spaces of manifolds. The developed methods apply to manifolds represented both parametrically and implicitly without preconditions such as knowledge of explicit formulas for geodesics. Hence, in addition to being interesting from a geometrical and computational point of view, the algorithms will be useful for applications in several of the mentioned areas.

To exemplify this, we consider the problem of capturing the variation of a set of manifold valued data. The well-known Principal Component Analysis procedure (PCA) has been generalized to manifold valued data with the introduction of Principal Geodesic Analysis (PGA, [10]). The construction is the source of continuing interest from both application oriented authors and the statistical community, most recently with the development of Geodesic PCA (GPCA, [18]). Both PGA and GPCA have been used successfully for a number of applications [10, 11, 18, 41, 35, 39].

Until now, there were no algorithm for numerically computing PGA for general manifolds. Linear approximations have been used instead except for special classes of manifolds where geodesics have explicit analytical formulas [35, 18]. Because PGA is posed as an optimization problem in the tangent space of the manifold, the tools developed here apply to computing it without linearizing the manifold. We will show how those tools allow us to compute exact PGA for a wide range of manifolds under some assumptions on the optimization problems.

1.1. Related Work. A vast body of mathematical literature describes manifolds and Riemannian structures, and [7, 26] provide excellent introductions to the field. Different aspects of numerical computation on implicitly defined manifolds are covered in [44, 34, 33]. Generalized inverses are important in the study of implicitly defined manifolds, and we will use a result of Decell [5].

An important starting point for our work is the paper of Dedieu and Nowicki [6] where the authors develop an initial value problem (IVP) for the computation of geodesics on implicitly defined manifolds. This result, together with the IVP defining geodesics in the parametrized case [7], constitutes the basis for the IVPs developed in the following sections. A similar approach is taken in [43] for computing Jacobi fields on the infinite dimensional manifold of diffeomorphisms. Several authors have studied the solution of the exponential map inverse problem, often called the logarithm map: in [29, 24, 36], different schemes are used to evolve an initial path towards a geodesic, and [22, 25, 39] use shooting methods. We build upon these works by assuming the logarithm problem is solved for the manifolds in question.

An optimization problem can be posed on a manifold in the sense that the domain of the cost function is restricted to the manifold and the sought for optima must reside on the manifold. Such problems are extensively covered in the literature (e.g. [28, 42]). The optimization problems we will solve involves the manifold geometry in the cost functions, but the domains will be the linear tangent spaces or subsets thereof with simple geometry. Therefore, the complexity will lie in the cost functions and not the optimization domains, and we will not need to use the optimization algorithms dealing with manifold domains.

The manifold generalization of linear PCA, PGA, was first introduced in [9], but it was formulated in the form most widely used in [10]. It has subsequently been used for several applications. To mention a few, the authors in [10, 11] study variations of medial atoms, [41] uses a variation of PGA for facial classification, [35] presents examples on motion capture data, and [39] applies PGA to vertebrae outlines. In addition, finding principal modes in tangent spaces, the procedure labeled linearized PGA in this paper, has been used for analyzing spine deformation modes and deformi-

ties in [3, 2]. The algorithm presented in [10] for computing PGA with tangent space linearization is most widely used. In contrast to this, [35] computes PGA as defined in [9] without approximations, but only for a specific manifold, the Lie group $SO(3)$. Our recent paper [38] uses the methods presented here to experimentally assess the effect of tangent space linearization, and we show that the algorithms work on high dimensional manifolds modelling real-life data.

A recent wave of interest in manifold valued statistics from the statistical community has lead to the development of Geodesic PCA (GPCA, [18, 19, 17]). GPCA is in many respects close to PGA but optimizes for the placement of the center point and minimizes projection residuals along geodesics instead of maximizing variance in geodesic subspaces. GPCA uses no linear approximation, but it is currently only computed on spaces where explicit formulas for geodesics exist and on quotients of such spaces.

1.2. Content and Outline. The paper will present the following main contributions:

- (1) We construct initial value problems allowing the computation of the differential of the exponential map and Jacobi fields, and second derivative of the exponential map on both parametric and implicitly represented manifolds of finite dimension.
- (2) We show how the tools developed allow for numerical computation of the sectional curvature and injectivity radius bounds for the manifolds.
- (3) We present an algorithm allowing the computation of PGA without linearizing the problem to the tangent space.
- (4) We present examples showing the differences between exact PGA and the linearized PGA previously used, and how the differences depend on the curvature of the manifold.

Due to the generality of the setup, the algorithm in (3) will work for many of the applications using PGA as defined in [10]. In particular, it will apply to those of the above mentioned examples using finite dimensional manifolds with available parametrization or implicit representation. We comment more on the classes of manifolds covered in section 2.1. In addition, we will need some assumptions on the manifold and dataset ensuring the optimization problems are well-behaved so that true global optima are found.

The importance of curvature computations is noted in [18], which lists the ability to compute sectional curvature as a high importance open problem. The result of (2) can be seen as a partial solution to this problem; we are indeed able to numerically compute the sectional curvature, although for the anomalous shape-spaces [23] used in [18] no parametrization or implicit representation is directly available, and hence the methods presented here do not apply.

In the experiments (4), we evaluate how the difference between the methods vary as we increase the curvature of the manifold. This experiment, which to the best of our knowledge has not been made before, is made possible by the generality of the algorithms of (1), which frees us from previous restrictions to specific manifolds such as $SO(3)$ [35] or anomalous shape-spaces [18].

The paper will start by a brief discussion of the required notation and geometry in section 2. We will touch upon the definition of PGA and how curvature and injectivity radius bounds relate to Jacobi fields. The reader already familiar with Riemannian geometry may wish to skip parts of this section. In section 3, we present IVPs for the differential of the exponential map and Jacobi fields and for the second derivative of

the exponential map. The actual derivations are lengthy and are, therefore, covered in the appendices. Following this, in section 4, we develop the exact PGA algorithm. We end the paper with experiments in section 5 and concluding remarks.

2. Geometry and Notation. We give a brief discussion of some aspects of differential and Riemannian geometry and, at the same time, introduce the notation used in the rest of the paper. The reader is referred to [7] for an introduction to differential geometry and Riemannian manifolds.

2.1. Manifolds and Their Representations. We will in the paper work with differentiable manifolds of finite dimension, and, in the sequel, M will denote such a manifold of dimension η . We will need M to be sufficiently smooth, i.e. of class C^k for $k = 3$ or 4 depending on the application. A chart of M is then a map $\varphi \in C^k(U, M)$ from an open subset U of \mathbb{R}^η to the manifold, and, since a chart provides a coordinate representation of a part of the manifold, it is often called a local parametrization.

Manifolds can be represented without local parameterizations. Let M be a level set of a differentiable map $F : \mathbb{R}^m \rightarrow \mathbb{R}^n$. If the Jacobian matrix $D_x F$ has full rank n for all $x \in M$, the level set is said to be regular. In that case, M will be an $(m-n)$ -dimensional manifold, and we say that M is *implicitly* defined. The space \mathbb{R}^m is called the embedding space. Throughout this paper, when dealing with implicitly defined manifolds, m and n will denote the dimension of the domain and codomain of F , respectively. We then have $\eta = m - n$ for the dimension η of the manifold,

In addition to local parametrizations and implicit representations, other ways of representing manifolds include discrete triangulations used for surfaces and quotients \tilde{M}/G of a larger manifold \tilde{M} by a group G . The latter is for example the case for Kendall's shape-spaces Σ_d^k [23]. Kendall's shape-spaces for planar points are actually complex projective spaces $\mathbb{C}P^{k-2}$ for which parameterizations are available, and, for points in 3-dimensional space and higher, the shape-spaces are anomalous and not manifolds. The spaces studied in [18] belong to this class.

Our methods do not apply directly to cases where local parametrizations or implicit representations are not available. We note, however, that for the quotients used in [18], \tilde{M} is a high-dimensional sphere and much of the optimization is performed on \tilde{M} instead of M/G . We are currently investigating how our methods can complement this in extending the approach to quotients \tilde{M}/G with \tilde{M} not restricted to being a sphere.

2.2. Curves and Differentiation. We will deal with parametrized entities, most notably curves on manifolds, and we use subscripts for the parameter. For example, a curve on M dependent on t will be denoted x_t . As our curves will normally start at $t = 0$, the starting point of curve x_t will be the point x_0 . The subscript notation should not be confused with differentiation with respect to the parameter t . When a local parametrization is available, we will often use it to represent the curve, and we will normally not distinguish between the curve and its expression $x_t = (x_t^1, \dots, x_t^\eta)$ in parameter space.

The tangent space of M at a point p is, a vector space of dimension η , will be denoted $T_p M$, and the derivative $\frac{d}{dt}x_t$ of a curve x_t evaluated at \tilde{t} then belongs to $T_{x_{\tilde{t}}} M$. We will often write just $\frac{d}{dt}x_{\tilde{t}}$ for such vectors, i.e. $\frac{d}{dt}x_t|_{t=\tilde{t}}$. In addition, when differentiating curves with respect to t , we often use the shorthand \dot{x}_t . With these conventions, $\frac{d}{dt}x_t|_{t=0}$, the initial velocity of the curve x_t , will be written \dot{x}_0 .

The differential of a map $f : M \rightarrow N$ will be denoted df and its evaluation at $p \in M$ will be denoted $d_p f$. When bases for $T_p M$ and $T_{f(p)} N$ are specified, or when

M and N are Euclidean spaces, we will write Df instead of df . We will encounter maps defined on a product of manifolds, e.g. $(v, w) \mapsto g(v, w) : M \times \tilde{M} \rightarrow N$, for which we will need to distinguish differentiation with respect to one of the variables only. Letting one of the parameters have a fixed value w_0 , the differential of the restricted function $v \mapsto g(v, w_0)$ from M to N evaluated at v_0 is denoted $d_{(v_0, w_0)}^v g$. Along the same lines, if V is a submanifold of M , the differential of $f|_V : V \rightarrow N$ will be denoted $d^{v \in V} f$ and its evaluation at $v_0 \in V$ will be written $d_{v_0}^{v \in V} f$.

2.3. Riemannian Manifolds and Geodesics. We will work solely with Riemannian manifolds, i.e. differentiable manifolds endowed with a smooth family of inner products on their tangent spaces. More precisely, a Riemannian metric on a manifold M is a smooth map g which associates to $p \in M$ an inner product $\langle \cdot, \cdot \rangle_p$ on $T_p M$, and, in a local parametrization, g will be a smooth map to the space of symmetric, positive definite matrices of order η . The pair (M, g) is then a Riemannian manifold. When M is a submanifold of \mathbb{R}^m , the tangent space $T_p M$ of M at a point p can be identified with a linear subspace of \mathbb{R}^m of dimension η , and the inner product $\langle \cdot, \cdot \rangle_p$ will be chosen to be the restriction of the standard inner product of \mathbb{R}^m .

The Riemannian metric determines notions such as length of curves, differentiation of vector fields, Christoffel symbols, and geodesics. If x_t is a curve, the length $l(x_t)$ is given by the integral $\int \|\dot{x}_t\| dt$ using the norm $\|\cdot\|$ on $T_{x_t} M$ induced by the metric. Computing directional derivatives of a vector field is done by a connection that associates to a pair (X, Y) of vector fields on M a new vector field denoted $\nabla_Y X$ so that $(\nabla_Y X)(p)$ will be a directional derivative of X at p in the direction $Y(p)$. A special connection, called the Levi-Civita connection, is associated to the Riemannian metric, and the connection defines the covariant derivative $\frac{D}{dt} V_t$ of a vector field V_t along a curve. On implicitly defined manifolds, $\frac{D}{dt} V_t$ is simply the projection of the usual derivative of vector fields onto $T_{x_t} M$, and, in a local parametrization, the covariant derivative of the vector fields $(\partial_{x_1}, \dots, \partial_{x_\eta})$ defines the Christoffel symbols Γ_{ij}^k of the metric by the relations $\nabla_{\partial_{x_i}} \partial_{x_j} = \sum_{k=1}^{\eta} \Gamma_{ij}^k \partial_{x_k}$. The η^3 functions $\Gamma_{ij}^k(x)$ satisfy the symmetry relation $\Gamma_{ij}^k = \Gamma_{ji}^k$.

Geodesic curves, manifold generalizations of straight lines, are characterized by having vanishing intrinsic acceleration expressed by the covariant derivative of the velocity field, $\frac{D}{dt} \dot{x}_t$ being zero. Geodesics are locally length minimizing and unique in the sense that given a point q and a velocity $v \in T_p M$, the geodesic passing q with velocity v is unique. The map which constructs geodesics given q and v is called the exponential map and denoted Exp . Thus, the unique geodesic is the curve $x_t = \text{Exp}_p t v$.

For points \tilde{q} in a sufficiently small neighborhood of q , the length minimizing curve joining q and \tilde{q} is unique as well. Given q and \tilde{q} , the initial direction in which to travel geodesically from q in order to reach \tilde{q} is given by the result of the logarithm map $\text{Log}_q(\tilde{q})$. We get the corresponding geodesic as the curve $t \mapsto \text{Exp}_q(t \text{Log}_q \tilde{q})$, and hence Log_q is the inverse of Exp_q . Subsets $\text{Exp}_q B_r(0)$ of M with $B_r(0)$ being a ball in $T_q M$ and with the radius $r > 0$ sufficiently small are examples of neighborhoods of q in which $\text{Log}_q(\tilde{q})$ is defined. Whenever we use the Log-map, we will restrict to such neighborhoods without explicitly mentioning it.

The gradient $\text{grad } h$ of a real valued function $h : M \rightarrow \mathbb{R}$ is also defined using the metric: at $p \in M$, $\text{grad}_p h$ is the unique vector in $T_p M$ which represents $d_p h$ in the sense that $d_p h(v) = \langle \text{grad}_p h, v \rangle$ for all $v \in T_p M$. Whenever a basis of $T_p M$ is specified, or when M is Euclidean, we switch to the usual notation ∇h . Similarly,

the Hessian of h is defined by the relation $\text{Hessian}(h)X = \nabla_X \text{grad } h$ for all vector fields X . Again, when a basis of $T_p M$ is specified, or when M is Euclidean, we use the usual notation $H(h)$.

2.4. Geodesic Systems. When a manifold is represented by a parametrization, the value of exponential map can be found as the solution of the IVP

$$\begin{aligned} \ddot{x}_t^k &= - \sum_{i,j}^{\eta} \Gamma_{ij}^k(x_t) \dot{x}_t^i \dot{x}_t^j, \quad k = 1, \dots, \eta \\ x_0 &= q, \quad \dot{x}_0 = v \end{aligned} \quad (2.1)$$

in parameter space at time $t = 1$. Recall that η denotes the dimension of the manifold and that a chart $\varphi : \mathbb{R}^\eta \rightarrow M$ is used to connect the parameter space and the manifold. This classical characterization of geodesics is not directly usable when the manifold is represented implicitly and, therefore, neither parametrization nor Christoffel symbols are directly available. To handle this situation, a first order IVP for the computation of the exponential map on implicitly represented manifolds as developed in [6]. Here $\text{Exp}_q v$ can be found as the x -part of the solution of the following IVP at time $t = 1$:

$$\begin{aligned} \dot{p}_t &= - \left(\sum_{k=1}^n \mu^k(x_t, p_t) H_{x_t}(F^k) \right) \dot{x}_t, \\ \dot{x}_t &= (I - D_{x_t} F^\dagger D_{x_t} F) p_t, \\ x_0 &= q, \quad p_0 = v. \end{aligned} \quad (2.2)$$

The map $\mu : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ is defined by $(x, p) \mapsto -(D_x F^T)^\dagger p$, and the symbol A^\dagger denotes the generalized inverse of the possibly non-square or singular matrix A [5].

2.5. Jacobi Fields and Global Geometry. Studying variations of geodesics leads to the notion of Jacobi fields, which encode important geometric information such as curvature and injectivity radius. In order to define Jacobi fields, let $x_{t,s}$ be a family of geodesics parametrized by s , i.e. for each \tilde{s} , the curve $t \mapsto x_{t,\tilde{s}}$ is a geodesic. When fixing the position t on the curves but varying the parameter s , we obtain the vector field $\frac{d}{ds} x_{t,0}$, and such a vector field is called a Jacobi field along the geodesic $x_{t,0}$.¹ The Jacobi fields along a given geodesic are uniquely determined by the initial conditions J_0 and $\frac{D}{dt} J_0$, the variation of the initial points $x_{0,s}$ and the covariant derivative of the field at $t = 0$, respectively. Define $q_s = x_{0,s}$, $v_s = \dot{x}_{0,s}$, and $w = \frac{d}{ds} v_0$. If $\frac{d}{ds} q_0 = J_0$ and $w = \frac{D}{dt} J_0$ then $\frac{d}{ds} \text{Exp}_{q_0}(tv_0)$ is equal to J_t [7, Chap. 5]. Therefore, in cases when q_s is constant and J_0 therefore 0, we have the following connection between J_t and $d\text{Exp}$:

$$d_{v_0} \text{Exp}_{q_0} tw = J_t. \quad (2.3)$$

Jacobi fields can equivalently be defined as solutions to the ODE

$$\frac{D^2}{dt^2} J_t = -R(\dot{x}_t, J_t) \dot{x}_t \quad (2.4)$$

with R denoting the curvature endomorphism [7, Chap. 5]. For parametrized manifolds, the ODE can be written in parameter space and can, in principle, be used for

¹ Recall that with the notation introduced in section 2.2, $\frac{d}{ds} x_{t,0}$ equals $\frac{d}{ds} x_{t,s}|_{s=0}$.

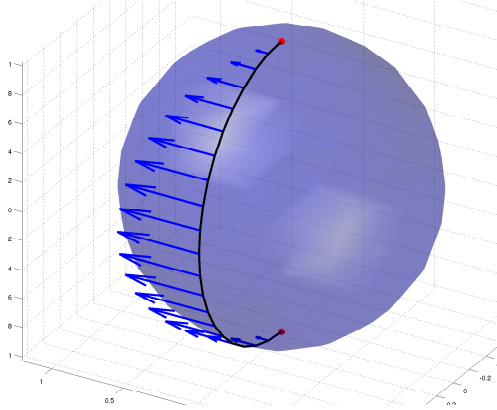


FIG. 2.1. The sphere \mathbb{S}^2 with a Jacobi field along a geodesic connecting the poles. Each pole is a conjugate point to the other since the non-zero Jacobi field vanishes. The injectivity radius is equal to the length of the geodesic, π .

numerical computations of Jacobi fields. The expressions are somewhat complicated, though, and we will obtain a different IVP by differentiating the system (2.1). The curvature endomorphism is not easily computed when the manifold is represented implicitly, and, therefore, the above ODE is not directly useful in this case. By differentiating the system (2.2), we remedy this in the next section.

Besides allowing us to calculate $d_{v_0} \text{Exp}_{q_0}$, Jacobi fields enable us to retrieve various geometric information about the manifold. We can for example estimate the sectional curvature of the manifold at q_0 using a Jacobi field J_t as defined above with $J_0 = 0$ and v_0, w orthonormal. Performing a Taylor expansion of the length $\|J_t\|$, we get

$$\|J_t\| = t - \frac{1}{6} K_{q_0}(\sigma) t^3 + O(t^4)$$

where $K_{q_0}(\sigma)$ is the sectional curvature of the plane span $\{v_0, w\}$ in $T_{q_0}M$ [7, Chap. 5]. For small t , the sectional curvature can then be estimated by

$$K_{q_0}(\sigma) \approx \frac{6}{t^3} (t - \|J(t)\|). \quad (2.5)$$

Furthermore, if J_t is a non-zero Jacobi field with $J_0 = 0$ along a geodesic x_t and, for some $\tilde{t} > 0$, also $J_{\tilde{t}} = 0$ then $x_{\tilde{t}}$ is called a conjugate point to x_0 . This implies that for any $r > \tilde{t}$, the geodesic x_t is not the shortest joining x_0 and x_r [7, Chap. 13]. In this way, we get an upper bound on the injectivity radius of M , which, in general terms, specifies the minimum length of non-minimizing geodesics. Figure 2.1 illustrates the situation on the sphere \mathbb{S}^2 .

2.6. Geodesic Subspaces. Linear subspaces are of great importance when studying data in Euclidean spaces; PCA, for example, can be formulated as an optimization problem on the set of linear subspaces. There is no obvious generalization of linear subspaces to manifolds, but, if one accepts the choice of a center point, the notion of geodesic subspaces becomes useful. A subset $\text{Exp}_q V$ of M is called a geodesic subspace centered at q if V is a linear subspace of $T_q M$. Geodesics between q and any point in the subspace are contained in the subspace, a fact which, in general, is not

true for geodesics between arbitrary pairs of points in the subspace. The projection of a point $x \in M$ onto a geodesic subspace $S = \text{Exp}_q V$ is defined as

$$\begin{aligned}\pi_S(x) &= \operatorname{argmin}_{y \in S} d(x, y)^2 = \operatorname{argmin}_{y \in S} \|\text{Log}_y x\|^2 \\ &= \text{Exp}_q(\operatorname{argmin}_{w \in V} \|\text{Log}_{\text{Exp}_q w} x\|^2) .\end{aligned}\quad (2.6)$$

Neither existence or uniqueness of the projection is in general ensured, although, for each geodesic subspace S , the set of points for which uniqueness fail has zero measure in M [18]. Existence of the projection is ensured if S is compact, which, for example, is the case if M is compact and S an embedded submanifold.

2.7. Principal Geodesic Analysis. Principal Component Analysis (PCA) is widely used to model the variability of data in Euclidean spaces. The procedure provides linear dimensionality reduction by defining a sequence of linear subspaces maximizing the variance of the projection of the data to the subspaces or, equivalently, minimizing the reconstruction errors. The k th subspace is spanned by an orthogonal basis $\{v^1, \dots, v^k\}$ of principal components v^1, \dots, v^k , and the i th principal component is defined recursively by

$$v^i = \operatorname{argmax}_{\|v\|=1} \frac{1}{N} \sum_{j=1}^N \left(\langle x_j, v \rangle^2 + \sum_{l=1}^{i-1} \langle x_j, v^l \rangle^2 \right) \quad (2.7)$$

when formulated as to maximize the variance of the projection of the dataset $\{x_1, \dots, x_N\}$ to the subspaces $\text{span}\{v^1, \dots, v^{i-1}\}$.

PCA is dependent on the vector space structure of the Euclidean space and hence cannot be performed on manifold valued datasets. Principal Geodesic Analysis was developed to overcome this limitation. PGA finds geodesic subspaces centered a point $\mu \in M$ with μ usually being an intrinsic mean² of the dataset $\{x_1, \dots, x_N\}$, $x_j \in M$. The k th geodesic subspace S_k of $T_\mu M$ is defined as $\text{Exp}_\mu(V_k)$ with $V_k = \text{span}\{v^1, \dots, v^k\}$ being the span of the principal directions v^1, \dots, v^k defined recursively by

$$\begin{aligned}v^i &= \operatorname{argmax}_{\|v\|=1, v \in V_{i-1}^\perp} \frac{1}{N} \sum_{j=1}^N d(\mu, \pi_{S_v}(x_j))^2 , \\ S_v &= \text{Exp}_\mu(\text{span}\{V_{i-1}, v\}) .\end{aligned}\quad (2.8*)$$

The notation V_{i-1}^\perp denotes the orthogonal complement of V_{i-1} in $T_\mu M$. The term being maximized is the sample variance of the projected data, the expected value of the squared distance to μ , and PGA therefore extends PCA by finding *geodesic subspaces* in which variance is maximized.

Since a method for computing the projection $\pi_{S_k}(x)$ has not been available for general manifolds, PGA has traditionally been computed using the orthogonal projection in the tangent space of μ to approximate the true projection. With this approximation, equation (2.8*) simplifies to

$$v^i \approx \operatorname{argmax}_{\|v\|=1} \frac{1}{N} \sum_{j=1}^N \left(\langle \text{Log}_\mu x_j, v \rangle^2 + \sum_{l=1}^{i-1} \langle \text{Log}_\mu x_j, v^l \rangle^2 \right)$$

² The notion of intrinsic mean goes back to Fréchet [14] and Karcher [21]. As in [10], we define it as $\operatorname{argmin}_{\mu \in M} \sum_{j=1}^N d(\mu, x_j)^2$. Uniqueness issues is treated in [21].

which is equivalent to (2.7), and, therefore, the procedure amounts to performing regular PCA on the vectors $\text{Log}_\mu x_j$. We will refer to PGA with the approximation as *linearized* PGA, and PGA as defined by (2.8*) will be referred to as *exact* PGA.

The above and prevalent definition of PGA is developed in [10], but a slightly different definition was introduced in [9]. The latter definition involves only one-dimensional subspaces and uses Lie group structure. In [35], the fact that π_S has a closed form solution on the sphere \mathbb{S}^3 when S is a one-dimensional geodesic subspace is used to compute exact PGA with the [9] definition by performing a steepest descent using the gradient of the cost function equivalent to the cost function of (2.8*).

Replacing maximization of sample variance by minimization of reconstruction error, we obtain another manifold extension of PCA and thus an alternate definition of PGA:

$$v^i = \operatorname{argmin}_{\|v\|=1, v \in V_{i-1}^\perp} \frac{1}{N} \sum_{j=1}^N d(x_j, \pi_{S_v}(x_j))^2. \quad (2.8**)$$

In contrast to vector space PCA, the two PGA definitions are *not* equivalent, a fact showing that the Euclidean and curved situations differ fundamentally. The latter formulation is chosen for Geodesic PCA to avoid instabilities of variance maximization [18], but the optimization algorithms developed in this paper work for both formulations. We will use the variance formulation for the experiments, but we will collectively refer to definitions by (2.8).

In general, PGA might not be well-defined as the mean might not be unique and both existence and uniqueness may fail for the projections (2.6) and the optimization problems (2.8). The convexity bounds of Karcher [21] ensures uniqueness of the mean for sufficiently local data, but setting up sufficient conditions to ensure well-posedness of (2.6) and (2.8) is a difficult issue, and here we will just assume well-posedness for the given manifold and dataset.

3. The Differentials. In this section, we aim at developing an initial value problem (IVPs) describing the differential of the exponential map and Jacobi fields, and, in addition, we will differentiate the IVPs a second time and thereby create the tools needed for the PGA algorithms presented in the next section. The basic strategy is simple: we differentiate the systems of section 2.4 and use the resulting IVPs.

It is a well-known fact that IVPs satisfying natural properties are differentiable with respect to their initial values [16, Chap. I.14]. The important contribution of this section is the explicit expressions for the differentiated systems that allow numerical integration and, in particular for the case of implicitly represented manifolds, are not straightforward to derive. To the best of our knowledge, no IVP describing the differential of the exponential map and Jacobi fields has previously been available in the implicit case; the IVP (3.3) remedies this situation. As previously noted, the ODE (2.4) describes Jacobi fields in the parameterized case but the expressions in parameter space are complicated. Therefore, we derive the IVP (3.2) below, which we find simpler to work with for the applications of this paper.

The presence of the generalized inverse in system (2.2) proves to be the main source of complexity for the implicit case. We handle the differentiation of this system using the following result of Decell:

THEOREM 3.1 ([5]). *Let A_s and its generalized inverse A_s^\dagger be differentiable s -dependent matrices. Then $\frac{d}{ds}(A_s^\dagger) = \Lambda(A_s, \frac{d}{ds}A_s)$ where*

$$\begin{aligned} \Lambda(A, B) = & -A^\dagger B A^\dagger + \left(B^T (A^\dagger)^T A^\dagger + A^\dagger (A^\dagger)^T B^T \right) \\ & - A^\dagger A \left(B^T (A^\dagger)^T A^\dagger + A^\dagger (A^\dagger)^T B^T \right) A A^\dagger. \end{aligned} \quad (3.1)$$

We will apply the result with $A_s = D_{x_{t,s}}F$ with $x_{t,s}$ an s dependent family of geodesics and t fixed. To see that $D_{x_{t,s}}F^\dagger$ is differentiable with respect to s when $x_{t,s}$ depends smoothly on s , take a frame of the normal space to M in a neighborhood of $x_{t,s}$, and note that $D_{x_{t,s}}F^\dagger$ is a composition of an invertible map onto the frame depending smoothly on s and the frame itself.

The remaining computations for deriving the systems are lengthy and notationally heavy. At this point, we only state the results and postpone the derivations and the proof of the following theorem to Appendix A.

THEOREM 3.2. *Let x_t be a geodesic in the C^3 manifold M with $x_0 = q$ and $\dot{x}_0 = v$, and let u, w be vectors in $T_{x_0}M$. Assume x_t is contained in a parametrized subset of M . Then the Jacobi field J_t along x_t with $J_0 = u$ and $\frac{D}{dt}J_0 = w$ can be found as the z -part of the solution of the IVP*

$$\begin{aligned} \begin{pmatrix} \dot{y}_t \\ \dot{z}_t \end{pmatrix} &= F_{q,v}^P \left(t, \begin{pmatrix} y_t \\ z_t \end{pmatrix} \right), \\ \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} &= \begin{pmatrix} w \\ u \end{pmatrix}, \end{aligned} \quad (3.2)$$

with $F_{q,v}^P$ the map given in explicit form in Appendix A.

Now, let instead $M \subset \mathbb{R}^m$ be defined as a regular zero level set of a C^3 map $F : \mathbb{R}^m \rightarrow \mathbb{R}^n$. Then the Jacobi field J_t along x_t with $J_0 = u$ and $\frac{D}{dt}J_0 = w$ can be found as the z -part of the solution of the IVP

$$\begin{aligned} \begin{pmatrix} \dot{y}_t \\ \dot{z}_t \end{pmatrix} &= F_{q,v}^I \left(t, \begin{pmatrix} y_t \\ z_t \end{pmatrix} \right), \\ \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} &= \begin{pmatrix} w \\ u \end{pmatrix}, \end{aligned} \quad (3.3)$$

with $F_{q,v}^I$ the map given in explicit form in Appendix A.

The maps $F_{q,v}^P$ and $F_{q,v}^I$ and consequently the systems (3.2) and (3.3) are linear in the initial values $(w \ u)^T$ as expected of systems describing differentials. They are non-autonomous due to the dependence on the position on the curve x_t .

The following corollary allows the computation of the derivative of the exponential map:

COROLLARY 3.3. *With the assumptions of Theorem 3.2, let (y_t, z_t) satisfy (3.2) or (3.3) with IVs $(w, 0)^T$. Then $d_v \text{Exp}_q w$ is equal to z_1 .*

Proof. Let J_t be the Jacobi field along x_t with $J_0 = 0$ and $\frac{D}{dt}J_0 = w$. By Theorem 3.2, $z_1 = J_1$, which, by (2.3), is equal to $d_v \text{Exp}_q w$.

□

The result enables us to compute the entire differential $d_v \text{Exp}_q$ by applying the corollary to each element of a basis $\{w^1, \dots, w^n\}$ for $T_q M$. The matrix having the results in its columns then equals $D_v \text{Exp}_q$. Note that $\text{Exp}_q \text{Log}_q y = y$ implies that

$d_y \text{Log}_q = (d_{\text{Log}_q} \text{Exp}_q)^{-1}$, a fact that allows the corollary to be used for computing $d_y \text{Log}_q$ as well.

We can differentiate the systems (3.2) and (3.3) once more if the manifold is sufficiently smooth. The main difficulty here is performing the algebra of the already complicated expressions for $F_{q,v}^P$ and $F_{q,v}^I$. For the implicit case, we will need to find the second derivative of $D_{x_{t,s}} F^\dagger$ and hence extend Decell's result. For simplicity, we consider a family of geodesics $x_{t,s}$ with the start point $x_{0,s}$ constant in s . The derivations and the proof are again postponed to Appendix A.

THEOREM 3.4. *Let $w \in T_q M$ with M of class C^4 , and let $x_{t,s}$ be a family of geodesics with $x_{0,s} = q$ and $v_s = \dot{x}_{0,s}$. Define $u = \frac{d}{ds} v_0$, and let $V_{q,v_0,w,u} = \frac{d}{ds} (d_{v_s} \text{Exp}_q w) = \frac{d}{ds} \left(\frac{d}{dr} (\text{Exp}_q v_s + r w) \right)$. Assume $x_{t,s}$ is contained in a parametrized subset of M . Then $V_{q,v_0,w,u}$ can be found as the r -part of the solution of the IVP*

$$\begin{aligned} \begin{pmatrix} \dot{q}_t \\ \dot{r}_t \end{pmatrix} &= G_{q,v_0,w,u}^P \left(t, \begin{pmatrix} q_t \\ r_t \end{pmatrix} \right), \\ \begin{pmatrix} q_0 \\ r_0 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \end{aligned} \tag{3.4}$$

with $G_{q,v_0,w,u}^P$ the map given in explicit form in Appendix A.

Now, let instead $M \subset \mathbb{R}^m$ be defined as a regular zero level set of a C^4 map $F : \mathbb{R}^m \rightarrow \mathbb{R}^n$. Then $V_{q,v_0,w,u}$ can be found as the r -part of the solution of the IVP

$$\begin{aligned} \begin{pmatrix} \dot{q}_t \\ \dot{r}_t \end{pmatrix} &= G_{q,v_0,w,u}^I \left(t, \begin{pmatrix} q_t \\ r_t \end{pmatrix} \right), \\ \begin{pmatrix} q_0 \\ r_0 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \end{aligned} \tag{3.5}$$

with $G_{q,v_0,w,u}^I$ the map given in explicit form in Appendix A.

We note that solutions to (3.4) and (3.5) depend linearly on u even though the systems are not linear.

3.1. Numerical Considerations. The geodesic systems (2.1) and (2.2) can in both the parametrized and implicit case be expressed in Hamiltonian forms. In [6], the authors use this property along with symplectic numerical integrators to ensure the computed curves will be close to actual geodesics. This is possible since the Hamiltonian encodes the Riemannian metric. Derivatives of Hamiltonian systems can be expressed in Hamiltonian form, and, therefore, the systems of Theorem 3.2 and Theorem 3.4 have Hamiltonian formulations. Using symplectic integrators, we can preserve the Hamiltonians, but the usefulness of this is limited since the Hamiltonians do not have directly interpretable forms in contrast to the case of geodesic systems.

Along the same lines, we would like to use the preservation of quadratic forms for symplectic integrators [15] to preserve quadratic properties of the differential of the exponential map, e.g. the Gauss Lemma [7]. At this point, we have, however, not been able to establish this for the implicit case.

4. Exact PGA. We will provide algorithms for iteratively solving the optimization problems (2.8) and hence compute exact PGA as defined in [10] without the traditional linear approximation. The algorithms will work for parametrized and implicitly represented manifolds under the following assumptions. First, we require that the PGA problem is well-defined as discussed in section 2.7. Second, the logarithm

map must be computable. As noted in the introduction, good implementations exist for both parametric and implicitly represented manifolds. Third, we will need to assume non-existence of local optima for the cost functions of (2.6) and (2.8) to ensure the optimization algorithms find the true global solutions. Forth, a local convexity assumption of the residual function, which is satisfied for local data, will be needed. We note that, if the third assumption is left out, it is indeed possible to find examples of manifolds and datasets where the algorithms will get stuck in local optima.

Solving the optimization problems (2.8) requires the ability to compute the projection operator π_S . We start by finding expressions for the gradients of the cost functions of the optimization problems using the IVPs derived in section 3, and, thereafter, we present the actual algorithms for solving the problems. The overall approach of solving (2.8) is similar to the approach of [35]. Our solution differs in that we are able to compute π_S and its differential without restricting to the manifold $SO(3)$ and in that we optimize (2.8) instead of the simpler³ cost function of [9].

The optimization problems (2.6) and (2.8) are posed in the tangent space of the manifold at the sample mean and the unit sphere of that tangent space, respectively. These domains have relatively simple geometry, and, therefore, the complexity of the problems is contained in the cost functions. Because of this, we will not need algorithms for optimizing problems with domains of complicated geometry.

As we are able to compute the gradient of the cost function of the problems, we can use approaches such as steepest descent. Yet, because both problems are quadratic, optimization algorithms such as Gauss-Newton or Levenberg-Marquardt are also applicable if the Jacobians are present. For simplicity, we compute gradients and present steepest descent algorithms, but it is straightforward to compute Jacobians instead and use more advanced optimization algorithms.

4.1. The Projection. We consider the projection $\pi_S(x)$ of a point $x \in M$ on a geodesic subspace S . Assume S is centered at $\mu \in M$, let V be a k -dimensional subspace of $T_\mu M$ such that $S = \text{Exp}_\mu V$, and define a residual function $R_{x,\mu} : V \rightarrow \mathbb{R}$ by $w \mapsto \|\text{Log}_{\text{Exp}_\mu w} x\|^2$ measuring distances between x and points in S . Computing $\pi_S(x)$ by solving (2.6) is then equivalent to finding $w \in V$ minimizing $R_{x,\mu}$. To find the gradient of $R_{x,\mu}$, choose an orthonormal basis for V and extend it to a basis for $T_\mu M$. Furthermore, let $w_0 \in V$ and choose an orthonormal basis for the tangent space $T_{\text{Exp}_\mu w_0} M$. Karcher showed in [21] that the gradient $\text{grad}^y \|\text{Log}_y x\|^2$ equals $-2\text{Log}_y x$, and, using this, we get the gradient of the residual function as

$$\nabla_{w_0}^{w \in V} R_{x,\mu} = -2(D_{w_0} \text{Exp}_\mu)^T_{1,\dots,k} (\text{Log}_{\text{Exp}_\mu w_0} x) \quad (4.1)$$

with $(D_{w_0} \text{Exp}_\mu)_{1,\dots,k}$ denoting the first k columns of $D_{w_0} \text{Exp}_\mu$ when expressed using the chosen bases.

4.2. The Gradient of the Projection. In order to optimize (2.8), we will need to compute gradients of the form

$$\text{grad}_{v_0}^{v \in V_{v_0}^\perp} d(y, \pi_{S_v}(x))^2 \quad (4.2)$$

with $V_{v_0} = \text{span}\{v^1, \dots, v^k, v_0\}$, $S_v = \text{Exp}(V_{v_0})$, and $y \in M$ being either the intrinsic mean μ for (2.8*) or x for (2.8**).⁴ This will involve the gradient of $\pi_{S_v}(x)$ with

³Simpler in the sense that projections in [9] involve only one-dimensional subspaces. The cost function of (2.8) uses i -dimensional subspaces for $i = 1, \dots, \eta$.

⁴ Since v in (2.8) is restricted to the unit sphere, we will not need the gradient in the direction of v_0 , and, therefore, we find the gradient in the subspace $V_{v_0}^\perp$ instead of in the larger space

respect to v . To derive this, we extend the domain of residual function $R_{x,\mu}$ defined in the previous subsection from V to $T_\mu M$. We will choose bases for $T_\mu M$ and V_{v_0} , and we let $H(R_{x,\mu})$ denote the Hessian of $R_{x,\mu}$ and $H(R_{x,\mu}|_{V_{v_0}})$ denote the Hessian of $R_{x,\mu}$ restricted to V_{v_0} with respect to the bases. Using this notation, we get the following result:

THEOREM 4.1. *Let $\{v^1, \dots, v^k\}$ be a basis for a subspace $V \subset T_\mu M$. For each $v \in V^\perp$, let V_v be the subspace $\text{span}\{V, v\}$, and let $S_v = \text{Exp}_\mu V_v$ be the corresponding geodesic subspace. Fix $v_0 \in V^\perp$ and define $w_0 = \text{Log}_\mu \pi_{S_{v_0}}(x)$ for an $x \in M$. Suppose the matrix $H_{v_0}(R_{x,\mu}|_{V_{v_0}})$ has full rank $k+1$. Extend the orthonormal basis $\{v^1, \dots, v^k, v_0/\|v_0\|\}$ for V_{v_0} to an orthonormal basis for $T_\mu M$. Then*

$$\begin{aligned} D_{v_0}^{v \in V_{v_0}^\perp} \pi_{S_v}(x) = & -(D_{w_0} \text{Exp}_\mu) \bar{v}_{x,\mu,v_0,S_{v_0}} \left(\nabla_{w_0}^{w \in V_{v_0}^\perp} R_{x,\mu} \right)^T \\ & + w_0^{k+1} (D_{w_0} \text{Exp}_\mu) E_{x,\mu,v_0,S_{v_0}}. \end{aligned} \quad (4.3)$$

The coordinates of the vector $\bar{v}_{x,\mu,v_0,S_{v_0}}$ in the basis for V_{v_0} are contained in the $(k+1)$ st column of the matrix $H_{v_0}(R_{x,\mu}|_{V_{v_0}})^{-1}$, the scalar w_0^{k+1} is the $(k+1)$ st coordinate of w_0 in the basis, and $E_{x,\mu,v_0,S_{v_0}}$ is the matrix

$$\begin{pmatrix} -H_{w_0}(R_{x,\mu}|_{V_{v_0}})^{-1} B_{w_0,v_0} \\ I_{\eta-(k+1)} \end{pmatrix}$$

with B_{w_0,v_0} the last $\eta - (k+1)$ columns of the matrix $(H_{w_0}(R_{x,\mu})(V \ v_0))^T$ and $I_{\eta-(k+1)}$ the identity matrix.

The proof of the theorem is presented in Appendix B. The assumption that the Hessian of the restricted residual $R_{x,\mu}|_{V_{v_0}}$ must have full rank is equivalent to the residual $R_{x,\mu}$ having only non-degenerate critical points when restricted to V_{v_0} . It is shown in [21] that $R_{x,\mu}$ is convex at points sufficiently close to x and the assumption is therefore satisfied in such cases. In order to compute the right hand side of (4.3), it is necessary to compute parts of the Hessian of the non-restricted residual $R_{x,\mu}$. The expression for computing $H_{v_0}(R_{x,\mu})$ is given in Appendix B.

Because $d(y, \pi_{S_v}(x))^2 = \|\text{Log}_y \pi_{S_v}(x)\|^2$, we have

$$\nabla_{v_0}^{v \in V_{v_0}^\perp} d(y, \pi_{S_v}(x))^2 = 2 \left((D_{\pi_{S_{v_0}}(x)} \text{Log}_y) (D_{v_0}^{v \in V_{v_0}^\perp} \pi_{S_v}(x)) \right)^T (\text{Log}_y \pi_{S_{v_0}}(x)), \quad (4.4)$$

which, combined with (4.3), gives (4.2).

4.3. Exact PGA Algorithm. The expressions for the gradients of the cost functions enable us to iteratively solve the optimization problems (2.6) and (2.8) under the mentioned assumptions. We let μ be the intrinsic mean of a dataset $\{x_1, \dots, x_N\}$ of points in M . The actual algorithms listed below are essentially steepest descent methods.

Algorithm 1 for computing $\pi_S(x)$ updates $w \in V$ instead of the actual point $y \in S$ that we are interested in. The vector w is related to y by $y = \text{Exp}_\mu w$.

For solving (2.8), we use that

$$\nabla_{v_0}^{v \in V_{v_0}^\perp} \left(\frac{1}{N} \sum_{j=1}^N d(y, \pi_{S_v}(x_j))^2 \right) = \frac{1}{N} \sum_{j=1}^N \nabla_{v_0}^{v \in V_{v_0}^\perp} d(y, \pi_{S_v}(x_j))^2. \quad (4.5)$$

$\text{span}\{v^1, \dots, v^k\}^\perp$.

Algorithm 1 Calculate $\pi_S(x)$

Require: $x \in M$, $S = \text{Exp}_\mu V$ geodesic subspace.

$w \leftarrow$ orthogonal projection of $\text{Log}_\mu x$ onto V {initial guess}

repeat

$y \leftarrow \text{Exp}_\mu w$ {vector to point}

$g \leftarrow -2(D_{w_0} \text{Exp}_\mu)^T_{1,\dots,k} \text{Log}_y x$ {gradient}

$\tilde{w} \leftarrow w$ {previous w }

$w \leftarrow w - g$ {update w }

until $\|\tilde{w} - w\|$ is sufficiently small.

Since v in (2.8) is required to be on the unit sphere, the optimization will take place on a manifold, and a natural approach to compute iteration updates will use the exponential map. Yet, because of the symmetric geometry of the sphere, we approximate this using the simpler method of adding the gradient to the previous guess and normalizing. When computing the $(k+1)$ st principal direction, we choose the initial guess as the first regular PCA vector of the data projected to V_k^\perp in $T_\mu M$. The algorithm for solving (2.8*) is listed in Algorithm 2, but by exchanging μ with x_j in the gradient computations and updating by subtracting the gradient, the algorithm will solve (2.8**) instead. See Figure 4.1 for an illustration of an iteration of the algorithm.

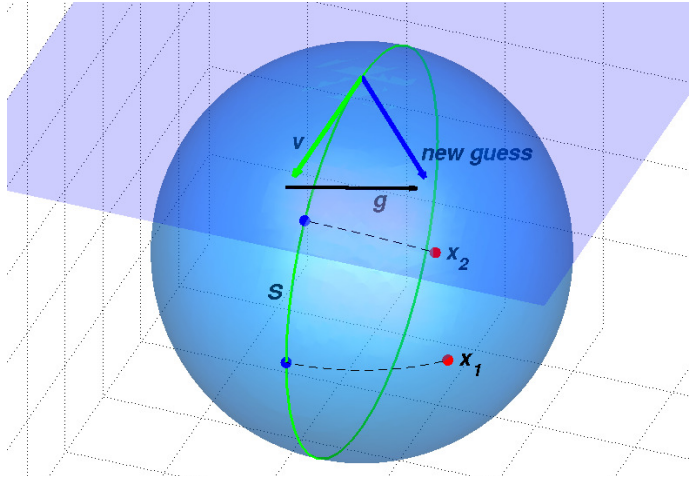


FIG. 4.1. An iteration of Algorithm 2. The figure shows data points x_1 and x_2 (red points) with projections (blue points) to the geodesic subspace S (green line). The vector v defining S is updated to the new guess by adding the gradient g .

5. Experiments. We will perform experiments exemplifying the differences between exact PGA and linearized PGA with synthetic data projected onto low dimensional manifolds on which it is possible to visually identify the differences between the methods. We vary the curvature of the manifolds in order to show how curvature affects the differences, and we compare the curvature approximation (2.5) and injectivity radius bound with the true values. For a comparison between the methods on high dimensional manifolds modelling real-life data, we refer the reader to [38]. In that paper, we compute and compare exact and linearized PGA on a 50 dimensional

Algorithm 2 Calculate the $(k + 1)$ st principal direction of (2.8*).

Require: $\mu, x_1, \dots, x_N \in M$, $\{v^1, \dots, v^k\}$ orthogonal basis for $V_k \subset T_\mu M$.

$v \leftarrow$ first PCA vector of $\{x_j\}$ projected first to $T_\mu M$
using Log_μ and then to V_k^\perp {initial guess}

repeat

$g_j \leftarrow \nabla_v^{v \in V_k^\perp} d(\mu, \pi_{S_v}(x_j))^2$ {for each j using (4.4)}

$g \leftarrow \frac{1}{N} \sum_{j=1}^N g_j$ {gradient using (4.5)}

$\tilde{v} \leftarrow v$ {previous v }

$v \leftarrow v + g$ {update v }

$v \leftarrow v / \|v\|$ {normalize}

until $\|\tilde{v} - v\|$ is sufficiently small.

manifold containing outlines of human vertebrae captured with lateral X-rays and on a 23 dimensional manifold containing human pose data acquired with tracking software.

The PGA algorithm is implemented in Matlab using Runge-Kutta ODE solvers. For the logarithm map, we use the shooting algorithm developed in [39]. All tolerances used for the integration and logarithm calculations are set at or lower than an order of magnitude of the precision used for the displayed results.

5.1. Synthetic Low-dimensional Data. We consider first surfaces embedded in \mathbb{R}^3 and defined by the equation

$$S_c = \{(x_1, x_2, x_3) | cx_1^2 + x_2^2 + x_3^2 = 1\}$$

for different values of the scalar c . For $c > 0$, S_c is an ellipsoid and it is equal to \mathbb{S}^2 in the case $c = 1$. The surface S_0 is a cylinder and, for $c < 0$, S_c is hyperboloid. Consider the point $p = (0, 0, 1)$ and note that $p \in S_c$ for all c . The curvature of S_c at p is equal to c . Note in particular that for the cylinder case the curvature is zero; the cylinder locally has the geometry of the plane \mathbb{R}^2 even though it informally seems to curve.

We evenly distribute 20 points along two straight lines through the origin of the tangent space $T_p S_c$, project the points from $T_p S_c$ to the surface S_c , and perform linearized and exact PGA using the variance formulation (2.8*). Figure 5.1 illustrates the situation in $T_p S_{-1}$ and on S_{-1} embedded in \mathbb{R}^3 , respectively.

Since linearized PCA amounts to Euclidean PCA in $T_p S_c$, the first principal direction found using linearized PGA divides the angle between the lines for all c . In contrast to this, the variance and the first principal direction found using exact PGA are dependent on c . Table 5.1 shows the angle between the principal directions found using the two methods, the variances and variance differences for different values of c .

c:	1	0.5	0	-0.5	-1	-1.5	-2	-3	-4	-5
angle ($^\circ$):	0.0	0.1	0.0	22.3	29.2	31.5	32.6	33.8	34.2	34.5
linearized var.:	0.899	0.785	0.601	0.504	0.459	0.435	0.423	0.413	0.413	0.417
exact var.:	0.899	0.785	0.601	0.525	0.517	0.512	0.510	0.508	0.507	0.506
difference:	0.000	0.000	0.000	0.212	0.058	0.077	0.087	0.095	0.094	0.089
difference (%):	0.0	0.0	0.0	4.2	12.5	17.6	20.6	23.0	22.7	21.4

TABLE 5.1

Differences between methods for different values of c .

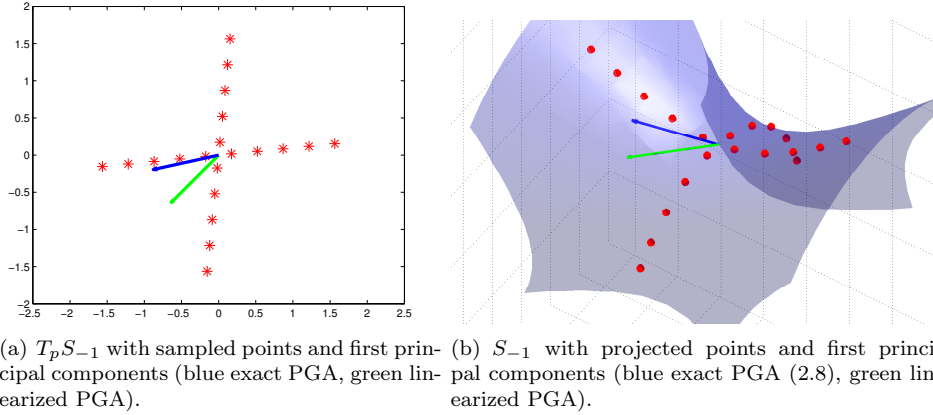


FIG. 5.1.

Let us give a brief explanation of the result. The symmetry of the sphere and the dataset cause the effect of curvature to even out in the spherical case S_1 . The cylinder S_0 has local geometry equal to \mathbb{R}^2 which causes the equality between the methods in the $c = 0$ case. The hyperboloids with $c < 0$, which can be constructed by revolving a hyperbola around its semi-minor axis, are non-symmetric causing an increase in variance as the first principal direction approaches the hyperbolic axis. The effect increases with the curvature causing the first principal direction to align with the hyperbolic axis for large negative values of c . We see that, for all negative values of c , exact PGA is able to capture more variance in the subspace spanned by the first principal direction than linearized PGA.

Using (2.5), we can approximate the sectional curvature K_p of S_c at p . The approximation is dependent on the value of the positive scalar t with increasing precision as t decreases to zero. Table 5.2 shows the result of the sectional curvature approximation for two values of t compared to the real curvature.

c :	1	0	-1	-2	-3
K_p :	1	0	-1	-2	-3
K_p est., $t = 0.01$:	1.000	0.000	-1.000	-2.000	-3.000
K_p est., $t = 0.1$:	1.000	0.000	-1.001	-2.002	-3.005

TABLE 5.2
Sectional curvature at p for different values of c .

Now let J_t be the Jacobi field with $J_0 = 0$ and $\frac{D}{dt}J_0 = (1, 0, 0)^T$ along the geodesic $x_t = \text{Exp}_p t(0, 1, 0)^T$. Figure 5.2 shows $\|J_t\|$ for different values of c . We see that $\|J_\pi\| = 0$ for the spherical case S_1 showing that x_1 is a conjugate point and hence giving the upper bound π on the injectivity radius. The situation is illustrated in Figure 2.1. The local geometric equivalence between the cylinder S_0 and \mathbb{R}^2 causes the straight line for $c = 0$. For all $c \leq 1$, the injectivity radius of S_c is π , but for $c < 1$, the point x_π not a conjugate point⁵. By looking at $\|J_t\|$, we are only able to detect conjugate points and hence, with this experiment, we only get the bound on the injectivity radius for $c \geq 1$. For $c > 1$ the injectivity radius decreases below 1 as seen in the case S_2 with $\|J_{\tilde{t}}\| = 0$ for $\tilde{t} \approx \pi/\sqrt{2}$.

⁵ For $c < 1$, x_π is a *cut* point [7, Chap. 13].

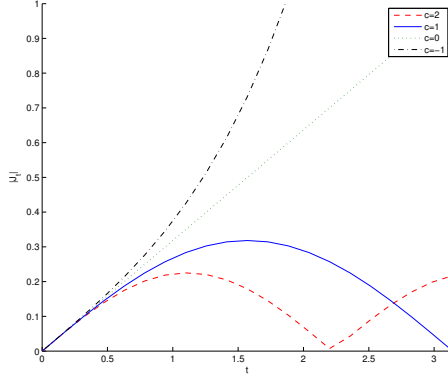


FIG. 5.2. $\|J_t\|$ for $c = 2, 1, 0, -1$ when $J_0 = 0$, $\frac{D}{dt}J_0 = (1, 0, 0)^T$, and $x_t = \text{Exp}_p t(0, 1, 0)^T$.

To investigate the difference with more than one principal direction, we consider a four dimensional manifold embedded in \mathbb{R}^5 and defined by

$$M_4 = \{(x_1, x_2, x_3, x_4, x_5) | x_1^2 - 2x_2^2 + x_3^2 - 2x_4 + x_5 = 1\}.$$

We make the situation more realistic than in the previous experiment by sampling 32 random points in the tangent space $T_p M_4$, $p = (0, 0, 0, 0, 1)$. Since $T_p M_4$ is an affine subspace of \mathbb{R}^5 orthogonal to the x_5 axis, we can identify it with \mathbb{R}^4 by the map $(x_1, x_2, x_3, x_4) \mapsto (x_1, x_2, x_3, x_4, 1)$. We use this identification when sampling by defining a normal distribution in \mathbb{R}^4 , sampling the 32 points from the distribution, and mapping the results to $T_p M_4$. The covariance is set to $\Sigma = \text{diag}(2, 1, 2/3, 1/3)$ to get non-spherical distribution and to increase the probability of data spreading over high-curvature parts of the manifold. Table 5.3 lists the variances and variance differences for the four principal directions for both methods along with angular differences. The lower variance for exact PGA compared to the linearized method for the 2nd principal direction is due to the greedy definition of PGA; when maximizing variance for the 2nd principal direction, we keep the first principal direction fixed. Hence we may get lower variance than what is obtainable if we were to maximize for both principal directions together.

Princ. comp.:	1	2	3	4
angle ($^\circ$):	10.1	10.6	12.0	12.2
linearized var.:	1.58	3.86	4.13	4.35
exact var.:	1.93	3.85	4.24	4.35
difference:	0.35	-0.01	0.11	0.00
difference (%):	21.9	-0.3	2.6	0.0

TABLE 5.3

Differences between the methods on M_4 . The variances of the data projected to the subspaces spanned by the first k principal directions and the percentage and angular differences are shown for $k = 1, \dots, 4$.

We clearly see angular differences between the principal directions. In addition, there is significant difference in accumulated variance in the first and third principal direction. We note that the percentage difference is calculated from what corresponds

to the accumulated spectrum. The percentage difference of the increase between the second and third principal direction, corresponding to the squared length of the third eigenvalue in regular PCA, is greater.

6. Conclusion. We have developed initial value problems allowing the computation of several important geometric structures on both parametrized and implicitly represented manifolds. We show how the constructed IVPs allow for numerical computation of injectivity radius bounds and sectional curvatures, which partially solves an open problem stated in [18]. Furthermore, the IVPs make possible computation of exact Principal Geodesic Analysis eliminating the need for the traditionally used linear approximations.

The experimental section presents examples of manifold valued datasets where exact PGA improves linearized PGA, and we show how the differences between the methods are dependent on the curvature of the manifolds. The differences are significant and clearly visually identifiable.

We are currently in the process of extending the methods to work for quotient manifolds M/G and thereby allowing the computations to be performed on practically all commonly occurring non-triangulated manifolds. We expect this would allow Geodesic PCA to be computed on general quotient manifolds as well. In addition, we are working on giving a theoretical treatment of the differences between the two formulations (2.8) of PGA. Finally, we expect to use the automatic computation of sectional curvatures to investigate further the effect of curvature on exact PGA and other statistical methods for manifold valued data.

Acknowledgements. The authors would like to thank P. Thomas Fletcher for fruitful discussions on how to compute exact PGA and Nicolas Courty for important remarks on problems linked to data locality. Furthermore, we wish to thank Søren Hauberg and Morten Engell-Nørregård for their help with producing data for testing the algorithms on high dimensional manifolds.

Appendix A. Derivation of the Derivative ODEs. We will use tensors on \mathbb{R}^n and \mathbb{R}^m for the proofs of Theorem 3.2 and Theorem 3.4, and we will use the common identification between tensors and multilinear maps, i.e. the tensor $T : (\mathbb{R}^k)^r \rightarrow \mathbb{R}$ defines a map multilinear map $\tilde{T} : (\mathbb{R}^k)^{r-1} \rightarrow \mathbb{R}^k$ by $\langle \tilde{T}(y_1, \dots, y_{r-1}), y_r \rangle = T(y_1, \dots, y_r)$. We will not distinguish between a tensor and its corresponding multilinear map, and hence, in the above case, write T for both maps.

For s -dependent vector fields $v_{s,1}, \dots, v_{s,r}$ and tensor field T_s , we will use the equality

$$\begin{aligned} \frac{d}{ds} T_0(v_{0,1}, \dots, v_{0,r}) \\ = \left(\frac{d}{ds} T_0 \right) (v_{0,1}, \dots, v_{0,r}) + T_0 \left(\frac{d}{ds} v_{0,1}, \dots, v_{0,r} \right) + \dots + T_0 \left(v_{0,1}, \dots, \frac{d}{ds} v_{0,r} \right) \end{aligned} \quad (\text{A.1})$$

for the derivative with respect to s . If T_{x_s} is a composition of an z -dependent tensor field T_z and an s -dependent curve x_s , the derivative $\frac{d}{ds} T_{x_s}$ equals the covariant tensor derivative $\nabla_{\frac{d}{ds} x_s} T_{x_s}$ [7, Chap. 4]. Since we will only use tensors on Euclidean spaces, such tensor derivatives will consist of component-wise derivatives.

In the following, we let T_z^P be the z -dependent 3-tensor on \mathbb{R}^n defined by

$$T_z^P(v_1, v_2, v_3) = - \sum_{i,j,k}^{\eta} \Gamma_{ij}^k(z) v_1^i v_2^j v_3^k$$

such that the k th component of $T_{x_t}^P(\dot{x}_t, \dot{x}_t)$ equals the right hand side of (2.1). Note that T_z^P is symmetric in the first two components since the Christoffel symbols are symmetric in i and j . Similarly, we let the z -dependent 3-tensor $T_z^{I,p}$ and 2-tensor $T_z^{I,x}$ equal the right hand side of the p and x parts of (2.2), respectively:

$$T_z^{I,p}(v_1, v_2) = - \left(\sum_{k=1}^n \mu^k(z, v_1) H_z(F^k) \right) v_2 ,$$

$$T_z^{I,x}(v) = (I - D_z F^\dagger D_z F) v$$

We carry out the proof of Theorem 3.2 in two parts starting with the parametrized case.

Proof. [Theorem 3.2] Let $x_{t,s}$ be a family of geodesics with $x_{t,0} = x_t$, and define $q_s = x_{0,s}$ and $v_s = \dot{x}_{0,s}$. Assuming $\frac{d}{ds} q_0 = u$ and $\frac{d}{ds} v_0 = w$, the Jacobi field J_t equals $\frac{d}{ds} \text{Exp}_{q_0}(tv_0)$, and, therefore, we can obtain J_t by differentiating the systems (2.1) and (2.2).

In the parametrized case, we get, using (A.1) and symmetry of T_z^P ,

$$\begin{aligned} \frac{d}{dt} \frac{d}{ds} x_{t,0} &= \frac{d}{ds} \ddot{x}_{t,0} = \frac{d}{ds} T_{x_{t,0}}^P(\dot{x}_{t,0}, \dot{x}_{t,0}) \\ &= \nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^P(\dot{x}_t, \dot{x}_t) + 2T_{x_{t,0}}^P\left(\frac{d}{dt} \frac{d}{ds} x_{t,0}, \dot{x}_t\right) , \\ \frac{d}{ds} x_{0,0} &= u, \quad \frac{d}{dt} \frac{d}{ds} x_{0,0} = w \end{aligned} \tag{A.2}$$

because $x_{t,s}$ are solutions to (2.1) with initial conditions q_s and v_s . Therefore, setting $y_t = \frac{d}{dt} \frac{d}{ds} x_{t,0}$ and $z_t = \frac{d}{ds} x_{t,0}$, we get (3.2) with

$$F_{q,v}^P(t, \begin{pmatrix} y_t \\ z_t \end{pmatrix}) = \begin{pmatrix} \nabla_{z_t} T_{x_t}^P(\dot{x}_t, \dot{x}_t) + 2T_{x_t}^P(y_t, \dot{x}_t) \\ y_t \end{pmatrix} .$$

As noted above, the derivative $\nabla_{\frac{d}{ds} x_{t,0}} T_{x_s}^P$ consists of just the component-wise derivatives of T_z^P , i.e. the derivatives of the Christoffel symbols.

For the implicit case, we use the map μ of section 2.4 to define the tensors

$$T_z^\mu(v) = \mu(z, v) , \quad T_z^H(v_1, v_2) = - \left(\sum_{k=1}^n v_1^k H_z(F^k) \right) v_2 ,$$

$$T_z^D(v) = (D_z F) v , \quad \text{and } T_z^{D^\dagger}(v) = (D_z F)^\dagger v .$$

Note, in particular, that $T_z^{I,p}(v_1, v_2) = T_z^H(T_z^\mu(v_1), v_2)$. We claim that $\frac{d}{ds} \text{Exp}_{q_0}(tv_0)$ equals the z -part of the solution of (3.3) with

$$\begin{aligned} &F_{q,v}^I \left(t, \begin{pmatrix} y_t \\ z_t \end{pmatrix} \right) \\ &= \begin{pmatrix} T_{x_t}^{I,p}(p_t, \dot{z}_t) + \nabla_{z_t} T_{x_t}^H(T_{x_t}^\mu(p_t), \dot{x}_t) + T_{x_t}^H(T_{x_t}^\mu(y_t) - \Lambda(T_{x_t}^D, \nabla_{z_t} T_{x_t}^D)^T p_t, \dot{x}_t) \\ T_{x_t}^{I,x}(y_t) - \Lambda(T_{x_t}^D, \nabla_{z_t} T_{x_t}^D) T_{x_t}^D(p_t) - T_{x_t}^{D^\dagger} \nabla_{z_t} T_{x_t}^D(p_t) \end{pmatrix} . \end{aligned} \tag{A.3}$$

Here $p_t = p_{t,0}$ where $p_{t,s}$ are the p -parts of the solutions to (2.2) with initial conditions q_s and v_s . To justify the claim, we differentiate the system (2.2). Using (A.1), we get

$$\begin{aligned} \frac{d}{dt} \frac{d}{ds} p_{t,0} &= \frac{d}{ds} \dot{p}_{t,0} = \frac{d}{ds} T_{x_{t,0}}^{I,p}(p_{t,0}, \dot{x}_{t,0}) \\ &= \nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^H(T_{x_t}^\mu(p_t), \dot{x}_t) + T_{x_t}^H(\nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^\mu(p_t) + T_{x_t}^\mu(\frac{d}{ds} p_{t,0}), \dot{x}_t) \\ &\quad + T_{x_t}^{I,p}(p_t, \frac{d}{ds} \dot{x}_{t,0}) \end{aligned}$$

and

$$\frac{d}{dt} \frac{d}{ds} x_{t,0} = \frac{d}{ds} \dot{x}_{t,0} = \frac{d}{ds} T_{t,0}^{I,x}(p_{t,0}) = \nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^{I,x}(p_t) + T_{x_t}^{I,x}(\frac{d}{ds} p_{t,0}) .$$

Note that the tensor derivative $\nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^H$ consists of derivatives of $H_{x_t}(F^k)$. Both the derivatives $\nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^\mu$ and $\nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^{I,x}$ involve derivatives of generalized inverses. Therefore, we apply Theorem 3.1 to differentiate $T_{x_t}^\mu$ and get that

$$\nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^\mu = -\Lambda(T_{x_t}^D, \nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^D)^T .$$

The tensor derivative $\nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^D$ consists of derivatives of $D_{x_t,s} F$. Similarly,

$$\nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^{I,x} = -\Lambda(T_{x_t}^D, \nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^D) T_{x_t}^D - T_{x_t}^{D^\dagger} \nabla_{\frac{d}{ds} x_{t,0}} T_{x_t}^D .$$

By differentiating the initial conditions, we get (3.3) with $y = \frac{d}{ds} p_{t,0}$, $z = \frac{d}{ds} x_{t,0}$, and $F_{q,v}^I$ as defined in (A.3). \square

For computing the second derivatives and proving Theorem 3.4, we will need to differentiate generalized inverses of matrices twice. For this task, we will use the lemma below, which follows directly from repeated application of the product rule for differentiation and Theorem 3.1.

LEMMA A.1. *Let $A_{t,s}$ be s - and t -dependent matrices. If $A_{t,s}$ and $A_{t,s}^\dagger$ are differentiable with respect to both variables and the mixed partial derivative $\frac{\partial^2}{\partial s \partial t} A_{t,s}$ exists, then*

$$\frac{\partial^2}{\partial s \partial t} (A_{t,s}^\dagger) = \tilde{\Lambda}(A_{t,s}, \frac{\partial}{\partial t} A_{t,s}, \frac{\partial}{\partial s} A_{t,s}, \frac{\partial^2}{\partial s \partial t} A_{t,s})$$

where

$$\begin{aligned} \tilde{\Lambda}(A, B, C, D) = & -\Lambda(A, C) B A^\dagger - A^\dagger D A^\dagger - A^\dagger B \Lambda(A, C) + Y(A, B, C, D) \\ & - \left(\Lambda(A, C) A + A^\dagger C \right) X(A, B) A A^\dagger - A^\dagger A Y(A, B, C, D) A A^\dagger \\ & - A^\dagger A X(A, B) \left(C A^\dagger + A \Lambda(A, C) \right) \end{aligned}$$

and

$$\begin{aligned} X(A, B) = & B^T (A^\dagger)^T A^\dagger + A^\dagger (A^\dagger)^T B^T , \\ Y(A, B, C, D) = & D^T (A^\dagger)^T A^\dagger + B^T \left(\Lambda(A, C)^T A^\dagger + (A^\dagger)^T \Lambda(A, C) \right) \\ & + \left(\Lambda(A, C) (A^\dagger)^T + A^\dagger (\Lambda(A, C))^T \right) B^T + A^\dagger (A^\dagger)^T D^T . \end{aligned}$$

We are then ready to prove Theorem 3.4. We will again start with the parameterized case, and we will use the tensors introduced in the beginning of this section and the proof of Theorem 3.2.

Proof. [Theorem 3.4] We compute the q and r parts of $G_{q,v_0,w,u}^P$ separately; denote them $G_{q,v_0,w,u}^{P,q}$ and $G_{q,v_0,w,u}^{P,r}$, respectively. Let $(y_{t,s}^w, z_{t,s}^w)$ be solutions to (3.2) with IV's $(w, 0)^T$ and along the geodesics $x_{t,s}$, and let y_t^w and z_t^w denote $y_{t,0}^w$ and $z_{t,0}^w$, respectively. Let also (y_t^u, z_t^u) be solutions to (3.2) with IV's $(u, 0)^T$ along $x_t = x_{t,0}$. Differentiating system (3.2), we get

$$\frac{d}{dt} \frac{d}{ds} (z_{t,0}^w) = \frac{d}{ds} (\dot{z}_{t,0}^w) = \frac{d}{ds} (y_{t,0}^w)$$

and, using symmetry of the tensors,

$$\begin{aligned} \frac{d}{dt} \frac{d}{ds} (y_{t,0}^w) &= \frac{d}{ds} (\dot{y}_{t,0}^w) = \frac{d}{ds} \nabla_{z_{t,0}^w} T_{x_{t,0}}^P (\dot{x}_{t,0}, \dot{x}_{t,0}) + 2 \frac{d}{ds} T_{x_{t,0}}^P (y_{t,0}^w, \dot{x}_{t,0}) \\ &= \nabla_{z_t^u} \nabla_{z_t^w} T_{x_t}^P (\dot{x}_t, \dot{x}_t) + \nabla_{\frac{d}{ds} z_{t,0}^w} T_{x_t}^P (\dot{x}_t, \dot{x}_t) + 2 \nabla_{z_t^w} T_{x_t}^P (y_t^u, \dot{x}_t) \\ &\quad + 2 \nabla_{z_t^u} T_{x_t}^P (y_t^w, \dot{x}_t) + 2 T_{x_t}^P (\frac{d}{ds} y_{t,0}^w, \dot{x}_t) + 2 T_{x_t}^P (y_t^w, y_t^u) . \end{aligned} \quad (\text{A.4})$$

Therefore, letting $q_t = \frac{d}{ds} y_{t,0}^w$ and $r_t = \frac{d}{ds} z_{t,0}^w$, we get $G_{q,v_0,w,u}^{P,q}(t, (r_t \ q_t)^T)$ as the right hand side of (A.4) and $G_{q,v_0,w,u}^{P,r}(t, (r_t \ q_t)^T)$ equal to q_t . The initial values are both 0 since $y_{0,s}^w$ and $z_{0,s}^w$ equal 0 and w , respectively, and, therefore, are not s -dependent.

For the implicit case, we will again compute the r and q parts of $G_{q,v_0,w,u}^I$ separately. Let now $(y_{t,s}^w, z_{t,s}^w)$ be solutions to (3.3) along the geodesics $x_{t,s}$ and with IV's $(w, 0)^T$, and let (y_t^u, z_t^u) be solutions to (3.3) along x_t and with IV's $(u, 0)^T$. Let also $p_{t,s}$ denote the p -parts of the solutions to (2.2) with initial conditions q and v_s , and write $p_t = p_{t,0}$, $y_t^w = y_{t,0}^w$, and $z_t^w = z_{t,0}^w$.

Differentiating system (3.3), we get

$$\begin{aligned} \frac{d}{dt} \frac{d}{ds} y_{t,0}^w &= \frac{d}{ds} \dot{y}_{t,0}^w = \frac{d}{ds} T_{x_{t,0}}^{I,p} (p_{t,0}, \dot{z}_{t,0}^w) + \frac{d}{ds} \nabla_{z_{t,0}^w} T_{x_{t,0}}^H (T_{x_{t,0}}^\mu (p_{t,0}), \dot{x}_{t,0}) \\ &\quad + \frac{d}{ds} T_{x_{t,0}}^H (T_{x_{t,0}}^\mu (y_{t,0}^w) - \Lambda(T_{x_{t,0}}^D, \nabla_{z_{t,0}^w} T_{x_{t,0}}^D)^T p_{t,0}, \dot{x}_{t,0}) . \end{aligned}$$

Using the map $\tilde{\Lambda}$ defined in Lemma A.1, we have

$$\frac{d}{ds} \Lambda(T_{x_{t,0}}^D, \nabla_{z_{t,0}^w} T_{x_{t,0}}^D)^T = \tilde{\Lambda}(T_{x_t}^D, \nabla_{z_t^w} T_{x_t}^D, \nabla_{z_t^u} T_{x_t}^D, \nabla_{z_t^u} \nabla_{z_t^w} T_{x_t}^D)^T .$$

Combining the equations, we get

$$\begin{aligned} \frac{d}{dt} \frac{d}{ds} y_{t,0}^w &= \nabla_{z_t^u} T_{x_t}^{I,p} (p_t, \dot{z}_t^w) + T_{x_t}^{I,p} (y_t^u, \dot{z}_t^w) + T_{x_t}^{I,p} (p_t, \frac{d}{dt} \frac{d}{ds} z_{t,0}^w) \\ &\quad + \nabla_{z_t^u} \nabla_{z_t^w} T_{x_t}^H (T_{x_t}^\mu (p_t), \dot{x}_t) + \nabla_{\frac{d}{ds} z_{t,0}^w} T_{x_t}^H (T_{x_t}^\mu (p_t), \dot{x}_t) \\ &\quad + \nabla_{z_t^w} T_{x_t}^H (T_{x_t}^\mu (y_t^u) - \Lambda(T_{x_t}^D, \nabla_{z_t^u} T_{x_t}^D)^T p_t, \dot{x}_t) + \nabla_{z_t^u} T_{x_t}^H (T_{x_t}^\mu (p_t), \dot{z}_t^u) \\ &\quad + \nabla_{z_t^u} T_{x_t}^H (T_{x_t}^\mu (y_t^w) - \Lambda(T_{x_t}^D, \nabla_{z_t^w} T_{x_t}^D)^T p_t, \dot{x}_t) \\ &\quad + T_{x_t}^H (T_{x_t}^\mu (\frac{d}{ds} y_{t,0}^w) - \Lambda(T_{x_t}^D, \nabla_{z_t^u} T_{x_t}^D)^T y_t^w, \dot{x}_t) \\ &\quad - T_{x_t}^H (\tilde{\Lambda}(T_{x_t}^D, \nabla_{z_t^u} T_{x_t}^D, \nabla_{z_t^u} \nabla_{z_t^w} T_{x_t}^D)^T p_t + \Lambda(T_{x_t}^D, \nabla_{z_t^u} T_{x_t}^D)^T y_t^u, \dot{x}_t) \\ &\quad + T_{x_t}^H (T_{x_t}^\mu (y_t^w) - \Lambda(T_{x_t}^D, \nabla_{z_t^w} T_{x_t}^D)^T p_t, \dot{z}_t^u) . \end{aligned}$$

Substituting $\frac{d}{ds} z_{t,0}^w$ with r_t and $\frac{d}{ds} y_{t,0}^w$ with q_t , we get $G_{q,v_0,w,u}^{I,q}$ as the right hand side of the equation. Likewise,

$$\begin{aligned} \frac{d}{dt} \frac{d}{ds} z_{t,0}^w &= \frac{d}{ds} T_{x_{t,0}}^{I,x} (y_{t,0}^w) - \frac{d}{ds} \Lambda(T_{x_{t,0}}^D, \nabla_{z_{t,0}^w} T_{x_{t,0}}^D)^T T_{x_{t,0}}^D (p_{t,0}) - \frac{d}{ds} T_{x_{t,0}}^{D^\dagger} \nabla_{z_{t,0}^w} T_{x_{t,0}}^D (p_{t,0}) \\ &= \nabla_{z_t^u} T_{x_t}^{I,x} (y_t^w) + T_{x_t}^{I,x} (\frac{d}{ds} y_{t,0}^w) \\ &\quad - \tilde{\Lambda}(T_{x_t}^D, \nabla_{z_t^u} T_{x_t}^D, \nabla_{z_t^u} \nabla_{z_t^w} T_{x_t}^D)^T T_{x_t}^D (p_t) \\ &\quad - \Lambda(T_{x_t}^D, \nabla_{z_t^u} T_{x_t}^D) \nabla_{z_t^u} T_{x_t}^D (p_t) - \Lambda(T_{x_t}^D, \nabla_{z_t^w} T_{x_t}^D) T_{x_t}^D (y_t^u) \\ &\quad - \Lambda(T_{x_t}^D, \nabla_{z_t^u} T_{x_t}^D) \nabla_{z_t^u} T_{x_t}^D (p_t) - T_{x_t}^{D^\dagger} \nabla_{z_t^u} \nabla_{z_t^w} T_{x_t}^D (p_t) - T_{x_t}^{D^\dagger} \nabla_{z_t^u} T_{x_t}^D (y_t^u) . \end{aligned}$$

Again, after substituting $\frac{d}{ds} y_{t,0}^w$ with q_t as above, we get $G_{q,v_0,w,u}^{I,r}$ as the right hand side of the equation. As for the parametric case, both initial values are zero.

□

Appendix B. The Projection Gradient. We prove Theorem 4.1, and, following this, we show how to compute the Hessian of the residual function $R_{x,\mu}$. We will need the following result for the proof of Theorem 4.1 to show that equation (4.3) is independent of the chosen basis.

LEMMA B.1. *Let S be an open subset of \mathbb{R}^k and $U : S \rightarrow M^{k \times (k-1)}$ a C^1 map with the property that for any $v \in S$, the columns of the matrix $(\frac{v}{\|v\|} U(v))$ constitute an orthonormal basis for \mathbb{R}^k . Let u_v^j denote the j th column of $U(v)$. Then for any $v_0 \in S$ and $w \in \mathbb{R}^k$, $\langle \frac{d}{dt} u_{v_0+t w}^j |_{t=0}, v_0 \rangle = -\langle u_{v_0}^j, w \rangle$. As consequence of this, if $\tilde{U} : S \rightarrow \mathbb{R}^{k-1}$ denotes the map $v \mapsto U(v)^T \frac{v_0}{\|v_0\|}$ then*

$$D_{v_0}^{v \in \text{span}(u_{v_0}^1, \dots, u_{v_0}^{k-1})} \tilde{U}(v) = -I_{k-1}$$

in the basis $u_{v_0}^1, \dots, u_{v_0}^{k-1}$ for $\text{span}(u_{v_0}^1, \dots, u_{v_0}^{k-1})$.

In the proof below, we adopt the notation of section 4.2, but we will use the alternative formulation $R_{x,\mu}(w) = \|\text{Log}_x \text{Exp}_\mu w\|^2$ for the residual function.

Proof. [Theorem 4.1] Extend the basis $\{v^1, \dots, v^k, v_0/\|v_0\|\}$ for V_{v_0} to an orthonormal basis for $T_\mu M$. The argument is not dependent on this choice of basis, but it will make the reasoning and notation easier. Let $S \subset T_\mu M \times V^\perp$ be an open neighborhood of (w_0, v_0) and define the map $F_V : S \rightarrow \mathbb{R}^\eta$ by

$$F_V(w, v) = \begin{pmatrix} \nabla_w R_{x,\mu} \cdot v^1 \\ \vdots \\ \nabla_w R_{x,\mu} \cdot v^k \\ \nabla_w R_{x,\mu} \cdot v \\ w \cdot u^1(v) \\ \vdots \\ w \cdot u^{\eta-k-1}(v) \end{pmatrix} = \begin{pmatrix} (V \ v)^T \nabla_w R_{x,\mu} \\ U_v^T w \end{pmatrix}$$

with the vectors $u^1(v), \dots, u^{\eta-(k+1)}(v)$ constituting an orthonormal basis for V_v^\perp for each v and with $(V \ v)$ and U_v denoting the matrices having v^i, v and $u^i(v)$ in the columns, respectively. Since $\langle \nabla_{w_0} R_{x,\mu}, v \rangle = d_{w_0} R_{x,\mu}(v) = 0$ for all $v \in V_{v_0}$ because w_0 is a minimizer for $R_{x,\mu}$ among vectors in V_{v_0} , we see that $F_V(w_0, v_0)$ vanishes. Therefore, if $D_{(w_0, v_0)}^w F_V$ is non-singular, the implicit function theorem asserts the existence of a map Ψ from a neighborhood of v_0 to $T_\mu M$ with the property that $F_V(\Psi(v), v) = 0$ for all v in the neighborhood. We then compute

$$0 = D_{v_0} F_V(\Psi(v), v) = \left(D_{(w_0, v_0)}^w F_V \right) (D_{v_0} \Psi(v)) + \left(D_{(w_0, v_0)}^v F_V \right)$$

and hence

$$D_{v_0}^{v \in V_{v_0}^\perp} \Psi(v) = - \left(D_{(w_0, v_0)}^w F_V \right)^{-1} \left(D_{(w_0, v_0)}^{v \in V_{v_0}^\perp} F_V \right). \quad (\text{B.1})$$

For the differentials on the right hand side of (B.1), we have

$$D_{(w_0, v_0)}^{v \in V_{v_0}^\perp} F_V = \left(0 \ \cdots \ 0 \ \nabla_{w_0}^{w \in V_{v_0}^\perp} R_{x,\mu} \ D_{(w_0, v_0)}^{v \in V_{v_0}^\perp} (w_0^T U_v) \right)^T$$

and

$$D_{(w_0, v_0)}^w F_V = \begin{pmatrix} (V \ v_0)^T & d_{w_0}^w(\nabla_w R_{x, \mu}) \\ U_{v_0}^T & \end{pmatrix} = \begin{pmatrix} (H_{w_0}(R_{x, \mu})(V \ v_0))^T \\ U_{v_0}^T \end{pmatrix}. \quad (\text{B.2})$$

With the choice of basis, the above matrix is block triangular,

$$D_{(w_0, v_0)}^w F_V = \begin{pmatrix} A_{w_0, v_0} & B_{w_0, v_0} \\ 0 & C_{w_0, v_0} \end{pmatrix}, \quad (\text{B.3})$$

with A_{w_0, v_0} equal to $H_{w_0}(R_{x, \mu}|_{V_{v_0}})$. The requirement that $D_{(w_0, v_0)}^w F_V$ is non-singular is fulfilled, because $H_{w_0}(R_{x, \mu}|_{V_{v_0}})$ has rank $k+1$ by assumption and U_{v_0} has rank $\eta - (k+1)$.

Since the first k rows of $D_{(w_0, v_0)}^{v \in V_{v_0}^\perp} F_V$ are zero, we need only the last $\eta - k$ columns of $(D_{(w_0, v_0)}^w F_V)^{-1}$ in order to compute (B.1). The vector $\bar{v}_{x, \mu, v_0, S_{v_0}}$ as defined in the statement of the theorem is equal to the $(k+1)$ st column. Let $E_{x, \mu, v_0, S_{v_0}}$ be the matrix consisting of the remaining $\eta - (k+1)$ columns. Using the form (B.3), we have

$$E_{x, \mu, v_0, S_{v_0}} = \begin{pmatrix} -H_{w_0}(R_{x, \mu}|_{V_{v_0}})^{-1} B_{w_0, v_0} C_{w_0, v_0}^{-1} \\ C_{w_0, v_0}^{-1} \end{pmatrix}.$$

Assume $\{u^1, \dots, u^j\}$ is chosen such that $\{u^1(v_0), \dots, u^j(v_0)\}$ equals the previously chosen basis for $V_{v_0}^\perp$. With this assumption, C_{w_0, v_0} is the identity matrix $I_{\eta - (k+1)}$. In addition, let w_0^{k+1} denote the $(k+1)$ st component of w_0 , that is, the projection of w_0 onto $v_0/\|v_0\|$. Since $w_0 \in V_{v_0}$ and by choice of U_v , Lemma B.1 gives

$$D_{(w_0, v_0)}^{v \in V_{v_0}^\perp} (U_v^T w) = w_0^{k+1} D_{(w_0, v_0)}^{v \in V_{v_0}^\perp} \left(U_v^T \frac{v_0}{\|v_0\|} \right) = -w_0^{k+1} I_{\eta - (k+1)}.$$

Therefore,

$$D_{(w_0, v_0)}^{v \in V_{v_0}^\perp} F_V = \begin{pmatrix} 0 & \dots & 0 & \nabla_{w_0}^{w \in V_{v_0}^\perp} R_{x, \mu} & -w_0^{k+1} I_{\eta - (k+1)} \end{pmatrix}^T.$$

Note, in particular, that $D_{(w_0, v_0)}^{v \in V_{v_0}^\perp} F_V$ is independent on the actual choice of bases U_v . Combining the equations, we get

$$D_{v_0}^{v \in V_{v_0}^\perp} \Psi(v) = -\bar{v}_{x, \mu, v_0, S_{v_0}} (\nabla_{w_0}^{w \in V_{v_0}^\perp} R_{x, \mu})^T + w_0^{k+1} E_{x, \mu, v_0, S_{v_0}}.$$

Because $\text{Exp}_\mu \Psi(v) = \pi_{S_v}(x)$, we get (4.3).

□

Lets now compute second derivatives, and thereby the Hessian, of the residual function $R_{x, \mu}$. Choose $w_0, v \in T_\mu M$ and let $y = \text{Exp}_\mu w_0$. Working in the orthonormal basis, we have

$$\nabla_{w_0} R_{x, \mu} = 2 \left((D_y \text{Log}_x) D_{w_0} \text{Exp}_\mu \right)^T \text{Log}_x y.$$

and hence

$$\begin{aligned} & \frac{d}{ds} (\nabla_{w_0 + vs} R_{x, \mu})|_{s=0} \\ &= 2 \left(\frac{d}{ds} \left(D_{\text{Exp}_\mu(w_0 + sv)} \text{Log}_x \right) |_{s=0} (D_{w_0} \text{Exp}_\mu) \right)^T \text{Log}_x y \\ & \quad + 2 \left((D_y \text{Log}_x) \frac{d}{ds} (D_{w_0 + vs} \text{Exp}_\mu) |_{s=0} \right)^T \text{Log}_x y \\ & \quad + 2 \left((D_y \text{Log}_x) (D_{w_0} \text{Exp}_\mu) \right)^T \frac{d}{ds} (\text{Log}_x \text{Exp}_\mu(w_0 + sv))|_{s=0}. \end{aligned} \quad (\text{B.4})$$

Note that

$$\frac{d}{ds} (\text{Log}_x \text{Exp}_\mu(w_0 + sv))|_{s=0} = (D_y \text{Log}_x) (D_{w_0} \text{Exp}_\mu) v .$$

Using that $\frac{d}{ds}(A_s^{-1}) = A_s^{-1}(\frac{d}{ds}A_s)A_s^{-1}$ for a time dependent, invertible matrix A_s ⁶ and the fact that $\text{Exp}_x \text{Log}_x z = z$ for all z , we get

$$\begin{aligned} \frac{d}{ds} \left(D_{\text{Exp}_\mu(w_0+sv)} \text{Log}_x \right) |_{s=0} &= \frac{d}{ds} \left(D_{\text{Log}_x(\text{Exp}_\mu w_0+sv)} \text{Exp}_x \right)^{-1} |_{s=0} \\ &= - (D_y \text{Log}_x) \frac{d}{ds} \left(D_{\text{Log}_x(\text{Exp}_\mu w_0+sv)} \text{Exp}_x \right) |_{s=0} (D_y \text{Log}_x) . \end{aligned}$$

The middle term of this product and the term $\frac{d}{ds} (D_{w_0+sv} \text{Exp}_\mu) |_{s=0}$ in (B.4) are both computable using Theorem 3.4.

REFERENCES

- [1] HARRY BLUM AND WEIANT WATHEN-DUNN, *A transformation for extracting new descriptors of shape*, Models for the Perception of Speech and Visual Form, (1967), pp. 380, 362.
- [2] JONATHAN BOISVERT, FARIDA CHERIET, XAVIER PENNEC, NICHOLAS AYACHE, AND HUBERT LABELLE, *Assessment of brace local action on vertebrae relative poses*, Studies in Health Technology and Informatics, 123 (2006), pp. 372–377. PMID: 17108454.
- [3] ———, *A novel framework for the 3D analysis of spine deformation modes*, Studies in Health Technology and Informatics, 123 (2006), pp. 176–181. PMID: 17108423.
- [4] VICENT CASELLES, RON KIMMEL, AND GUILLERMO SAPIRO, *Geodesic active contours*, International Journal of Computer Vision, 22 (1995), pp. 61–79.
- [5] HENRY P. DECELL, *On the derivative of the generalized inverse of a matrix*, Linear and Multilinear Algebra, 1 (1974), p. 357.
- [6] JEAN-PIERRE DEDIEU AND DMITRY NOWICKI, *Symplectic methods for the approximation of the exponential map and the newton iteration on riemannian submanifolds*, Journal of Complexity, 21 (2005), pp. 487–501.
- [7] MANFREDO PERDIGAO DO CARMO, *Riemannian geometry*, Mathematics: Theory & Applications, Birkhauser Boston Inc., Boston, MA, 1992.
- [8] RICARDO FERREIRA, JOAO XAVIER, JOAOA COSTERIA, AND VICTOR BARROSO, *Newton algorithms for riemannian distance related problems on connected locally symmetric manifolds*, Technical Report, Signal and Image Processing Group (SPIG), Institute for Systems and Robotics (ISR), (2008).
- [9] P.T. FLETCHER, CONGLIN LU, AND S. JOSHI, *Statistics of shape via principal geodesic analysis on lie groups*, in CVPR 2003, vol. 1, 2003, pp. I–95–I–101 vol.1.
- [10] P.T. FLETCHER, CONGLIN LU, S.M. PIZER, AND SARANG JOSHI, *Principal geodesic analysis for the study of nonlinear statistics of shape*, Medical Imaging, IEEE Transactions on, 23 (2004), pp. 995–1005.
- [11] P. THOMAS FLETCHER AND SARANG JOSHI, *Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors*, ECCV Workshops CVAMIA and MMBIA., 3117 (2004), pp. 87–98.
- [12] ———, *Riemannian geometry for the statistical analysis of diffusion tensor data*, Signal Processing, 87 (2007), pp. 250–262.
- [13] P. THOMAS FLETCHER, SURESH VENKATASUBRAMANIAN, AND SARANG JOSHI, *Robust statistics on riemannian manifolds via the geometric median*, in 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 2008, pp. 1–8.
- [14] M. FRÉCHET, *Les éléments aléatoires de nature quelconque dans un espace distancié*, Ann. Inst. H. Poincaré, 10 (1948), pp. 215–310.
- [15] ERNST HAIRER, CHRISTIAN LUBICH, AND GERHARD WANNER, *Geometric numerical integration*, Springer, 2002.
- [16] ERNST HAIRER, SYVERT P. NØRSETT, AND GERHARD WANNER, *Solving Ordinary Differential Equations I: Nonstiff Problems (Springer Series in Computational Mathematics)*, Springer, 2nd ed., May 2008.

⁶ See [5, Eq. (2)].

- [17] STEPHAN HUCKEMANN AND THOMAS HOTZ, *Principal component geodesics for planar shape spaces*, Journal of Multivariate Analysis, 100 (2009), pp. 699–714.
- [18] STEPHAN HUCKEMANN, THOMAS HOTZ, AND AXEL MUNK, *Intrinsic shape analysis: Geodesic PCA for riemannian manifolds modulo isometric lie group actions*, Statistica Sinica, 20 (2010), pp. 1–100.
- [19] STEPHAN HUCKEMANN AND HERBERT ZIEZOLD, *Principal component analysis for riemannian manifolds, with an application to triangular shape spaces*, Advances in Applied Probability, 38 (2006), pp. 299–319.
- [20] SARANG JOSHI, STEPHEN PIZER, P THOMAS FLETCHER, PAUL YUSHKEVICH, ANDREW THALL, AND J S MARRON, *Multiscale deformable model segmentation and statistical shape analysis using medial descriptions*, IEEE Transactions on Medical Imaging, 21 (2002), pp. 538–550. PMID: 12071624.
- [21] H. KARCHER, *Riemannian center of mass and mollifier smoothing*, Communications on Pure and Applied Mathematics, 30 (1977), pp. 509–541.
- [22] HERBERT BISHOP KELLER, *Numerical methods for two-point boundary-value problems*, Blaisdell, (Waltham, Mass), 1968.
- [23] DAVID G. KENDALL, *Shape manifolds, procrustean metrics, and complex projective spaces*, Bull. London Math. Soc., 16 (1984), pp. 81–121.
- [24] ERIC KLASSEN AND ANUJ SRIVASTAVA, *Geodesics between 3D closed curves using Path-Straightening*, in ECCV 2006, vol. 3951, Springer, 2006, pp. 95–106.
- [25] ERIC KLASSEN, ANUJ SRIVASTAVA, WASHINGTON MIO, AND SHANTANU JOSHI, *Analysis of planar shapes using geodesic paths on shape spaces*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26 (2004), pp. 372–383.
- [26] JOHN M LEE, *Riemannian manifolds*, vol. 176 of Graduate Texts in Mathematics, Springer-Verlag, New York, 1997. An introduction to curvature.
- [27] ZHENG DONG LU, MIGUEL CARREIRA-PERPINAN, AND CHRISTIAN SMINCHISESCU, *People tracking with the laplacian eigenmaps latent variable model*, in Advances in Neural Information Processing Systems 20, MIT Press, 2008, pp. 1705–1712.
- [28] DAVID G. LUENBERGER, *The gradient projection method along geodesics*, Management Science, 18 (1972), pp. 620–631. ArticleType: primary_article / Issue Title: Theory Series / Full publication date: Jul., 1972 / Copyright © 1972 INFORMS.
- [29] LYLE NOAKES, *A global algorithm for geodesics*, Journal of the Australian Mathematical Society, 64 (1998), pp. 37–50.
- [30] XAVIER PENNEC, *Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements*, J. Math. Imaging Vis., 25 (2006), pp. 127–154.
- [31] XAVIER PENNEC, PIERRE FILLARD, AND NICHOLAS AYACHE, *A riemannian framework for tensor computing*, Int. J. Comput. Vision, 66 (2006), pp. 41–66.
- [32] XAVIER PENNEC, CHARLES GUTTMANN, AND JEAN-PHILIPPE THIRION, *Feature-based registration of medical images: Estimation and validation of the pose accuracy*, in MICCAI 1998, Springer Berlin, 1998, pp. 1107–1114.
- [33] PATRICK J. RABIER AND WERNER C. RHEINOLDT, *On a computational method for the second fundamental tensor and its application to bifurcation problems*, Numerische Mathematik, 57 (1990), pp. 681–694.
- [34] W. C. RHEINOLDT, *MANPAK: a set of algorithms for computations on implicitly defined manifolds*, Computers & Mathematics with Applications, 32 (1996), pp. 15–28.
- [35] SALEM SAID, NICOLAS COURT, NICOLAS LE BIHAN, AND STEPHEN SANGWINE, *Exact principal geodesic analysis for data on $so(3)$* , EUSIPCO 2007, (2007).
- [36] FRANK SCHMIDT, MICHAEL CLAUSEN, AND DANIEL CREMERS, *Shape matching by variational computation of geodesics on a manifold*, in Pattern Recognition, Springer Berlin, 2006, pp. 142–151.
- [37] CRISTIAN SMINCHISESCU AND ALLAN JEPSON, *Generative modeling for continuous Non-Linearly embedded visual inference*, In ICML, (2004), pp. 759–766.
- [38] STEFAN SOMMER, FRANCOIS LAUZE, SØREN HAUBERG, AND MADS NIELSEN, *Manifold valued statistics, exact principal geodesic analysis and the effect of linear approximations*, in ECCV 2010, vol. 6316 of Lecture Notes in Computer Science, Heraklion, Greece, 2010, Springer, Heidelberg, pp. 43–56.
- [39] STEFAN SOMMER, ADITYA TATU, CHEN CHEN, DAN JØRGENSEN, MARLEEN DE BRUIJNE, MARCO LOOG, MADS NIELSEN, AND FRANCOIS LAUZE, *Bicycle chain shape models*, in MM-BIA/CVPR 2009, 2009, pp. 157–163.
- [40] RAQUEL URTASUN, DAVID J. FLEET, AARON HERTZMANN, AND PASCAL FUA, *Priors for people tracking from small training sets*, in 2005 IEEE International Conference on Computer Vision (ICCV), IEEE Computer Society, 2005, pp. 403–410.

- [41] JING WU, W. SMITH, AND E. HANCOCK, *Weighted principal geodesic analysis for facial gender classification*, in Progress in Pattern Recognition, Image Analysis and Applications, Springer Berlin, 2008, pp. 331–339.
- [42] Y. YANG, *Globally convergent optimization algorithms on riemannian manifolds: Uniform framework for unconstrained and constrained optimization*, Journal of Optimization Theory and Applications, 132 (2007), pp. 245–265.
- [43] LAURENT YOUNES, FELIPE ARRATE, AND MICHAEL I. MILLER, *Evolutions equations in computational anatomy*, NeuroImage, 45 (2009), pp. S40–S50.
- [44] QIN ZHANG AND GUOLIANG XU, *Curvature computations for n -manifolds in and solution to an open problem proposed by r. goldman*, Computer Aided Geometric Design, 24 (2007), pp. 117–123.